AD-A110 955 MINNESOTA UNIV MINNEAPOLIS DEPT OF PSYCHOLOGY F/G 5/10
DIMENSIONALITY OF MEASURED ACHIEVEMENT OVER TIME.(U)
DEC 81 K A GIALLUCA, D J WEISS N00014-79-C-0172
UNCLASSIFIED RR-81-5 NL

1 OF 1
AD
A110955

END
DATE
FILMED
3-82
DTIC

LEVEL II

(12)

Dimensionality of Measured Achievement Over Time

AD A110955

Kathleen A. Gialluca
and
David J. Weiss

RESEARCH REPORT 81-5
DECEMBER 1981

COMPUTERIZED ADAPTIVE TESTING LABORATORY
PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Air Force Human Resources Laboratory, the Army Research Institute, the Air Force Office of Scientific Research, and the Office of Naval Research, and monitored by the Office of Naval Research.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 81-5	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Dimensionality of Measured Achievement Over Time		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kathleen A. Gialluca and David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0172
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, Minnesota 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 6115N Proj.: RR042-04 T.A.: RR042-04-01 W.U.: NR 150-433
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE December, 1981
		13. NUMBER OF PAGES 34
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This research was supported by funds from the Air Force Human Resources Laboratory, the Army Research Institute, the Air Force Office of Scientific Research, and the Office of Naval Research, and monitored by the Office of Naval Research.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

achievement testing
change scores
measurement of change
measurement of growth
factor comparisons
factor structure

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Some type of difference or change score is frequently used to quantify the effects of experimental treatments and educational programs on individuals and on groups of individuals. Whether the change measurement involves the use of simple difference scores, their derivatives, or some more complex methodological design, the measurement process itself assumes that the treatment or instruction results in higher levels of the originally measured variable and that the only change that occurs is a quantitative one. If this assumption is

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

not met, then the computation of any type of difference score is inappropriate and the scores themselves are useless for measuring growth or change.

Two studies investigated the tenability of the assumption that classroom instruction results in increases in students' achievement levels while the qualitative nature of that achievement remains constant across time. The data utilized were the item responses to tests in basic mathematics and in general biology administered as pretests and after instruction to students enrolled in those courses.

Results indicated that this assumption was not tenable in the biology data set, where increases in mean achievement level were accompanied by corresponding changes in the factor structure underlying the item responses. For the mathematics data, however, there was no such violation of the assumption: As student achievement levels increased the underlying factor structure remained unchanged. The implications of these results for psychology, education, and program evaluation are noted.

Contents

Acknowledgments

Data utilized in Study I of this report were obtained from students enrolled in General College mathematics courses at the University of Minnesota during fall quarter 1979. Appreciation is extended to these students and to Douglas Robertson, Mathematics Coordinator of General College, for their participation in this research.

Data utilized in Study II of this report were obtained from volunteer students in General Biology, Biology 1-011, at the University of Minnesota during winter quarter 1980. Appreciation is extended to these students, and to Kathy Swart and Norman Kerr of the General Biology staff, for their participation in this research. Gage Kingsbury and Elana Broch were responsible for the research design and collection of biology data during Winter 1980, of which the data reported herein were a part.

Technical Editor: Barbara Leslie Camm

DIMENSIONALITY OF MEASURED ACHIEVEMENT OVER TIME

The measurement of individual or group change is central to many issues in the fields of psychology, education, and program evaluation. Psychologists, educators, and (more recently) evaluators typically use differences in test scores to quantify the effects of experimental treatments and educational programs on individuals and on groups of individuals.

The typical paradigm for measuring change involves the administration of a standardized achievement test both before and after an experimental treatment or program implementation; the effect of the treatment intervention is then considered to be a function of the mean difference between the two sets of test scores. If two or more groups of students are involved, comparisons can also be made between treatment and control groups, or among groups exposed to various treatments or involved in several different programs. Again, evaluation of treatment effects involves comparing the mean achievement gain (typically, a function of the difference scores) observed for each group. Individual gain or change is also frequently used to measure an individual's growth in achievement level or change due to a treatment or special program.

Lord (1963) and Cronbach and Furby (1970), among others, have discussed the methodological and statistical problems involved in using difference scores to measure change or growth and have presented some possible solutions. Whether measurements of change involve the use of simple difference scores, their derivatives, or some more complex methodological design, the measurement process itself assumes that the treatment or instruction results in increased levels of the same trait or characteristic that was measured originally and that the only change that occurs is a quantitative one.

That this assumption may be violated has long been evident in studies of intelligence and intellectual growth. Garrett (1946) noted that "intelligence changes in its organization" (p. 373) and called for corresponding changes in the way intelligence is measured. This "differentiation hypothesis" spawned much research (see Reinert, 1970, for a review) concerning the changes in the structure and organization of intelligence throughout the human life span. Some of these studies report results supporting the hypothesis of age differentiation; others offer support for a hypothesis of age integration, and still others provide evidence in support of both these hypotheses. Nearly all this research, however, has found that the structure of intelligence, as defined by factor analysis, does not remain constant with age and experience.

Other authors (Anastasi, 1936; Ferguson, 1954; Games, 1962; Woodrow, 1938, 1939a, 1939b, 1939c) have investigated the changes in verbal ability and intellectual factor structure that accompany shorter term training and practice. Similar factor-analytic investigations have been made in the areas of psychomotor behavior (Fleishman, 1953, 1957, 1960; Fleishman & Hempel, 1954, 1955; Greene, 1943), psycholinguistic abilities (Querishi, 1967), word association (Sullivan & Moran, 1967; Swartz & Moran, 1968), and even the learning of Morse code (Fleishman & Fruchter, 1960). All these authors have found that the facto-

rial structure of abilities underlying task performance changes in a systematic way with training and practice. An individual's status at a later point in time, then, may be qualitatively different from his/her status as originally measured.

Wohlwill (1970) discusses this issue of quantitative versus qualitative change more generally in the area of developmental psychology and, like Garrett (1946), calls for more sophisticated scaling methods which will

> ... allow us to assess an individual's status on a developmental dimension in a manner such as to ensure not only comparability of content for the different parts of that dimension, but at the same time a continuous scale along which developmental change can be charted .... Postulating a unitary dimension across the age span under investigation presupposes that there are no major discontinuities in the development of the behavior in question, such as there obviously are in the assessment of intelligence when we move from infancy to childhood. (p. 154)

Although Reinert (1970) called for the investigation of possible factor-structure changes in areas other than intelligence and abilities more than a decade ago, no research has yet extended this line of questioning into the area of classroom achievement. That is, there have been no reported studies that have systematically investigated whether the individual and group changes that occur after classroom instruction or program participation are quantitative changes in the level of achievement, as is generally assumed, or whether more qualitative changes in the structure of the achievement variable have occurred.

Kingsbury and Weiss (1979) studied the effects of testing students at different points in instruction. They reported that the single factor extracted from the item responses to a college general biology examination administered on the first day of class and the factor extracted from the item responses to a classroom midquarter examination differed markedly from each other in terms of strength; however, they could not further investigate the similarity of the factor pattern loadings from both administrations. They cautioned that replications of their findings contrasting the pretest factor with the later achievement factor would render difference scores "completely useless" as indicators of achievement level growth, since different variables would, in fact, be measured at the two points in time.

The importance of such a conclusion should not be underestimated. If different characteristics are, in fact, being measured at two different occasions, then the computation of any type of difference score is inappropriate, and the evaluation of program effectiveness and gains in individual student achievement must be made on some other basis. It is justifiable to use difference scores (statistical and methodological issues notwithstanding) only when it can be demonstrated that quantitative changes are the only changes accompanying instruction.

Purpose

The objectives of the present studies were to investigate the nature of the changes in the dimensionality of achievement that occurred following instruction in two different achievement domains--basic mathematics and general biology--and

to determine the appropriateness of calculating difference scores in order to measure change in these domains.

STUDY I

Method

Subjects and Tests

Data were obtained from students enrolled in mathematics classes at the University of Minnesota's General College during the fall quarter of 1979. These students were administered a 35-item Arithmetic Placement Test (APT) on the first day of class (pretest) and again as a final examination (posttest). The APT is composed of five-alternative multiple-choice items covering such topics as addition, subtraction, multiplication, and division of whole numbers, fractions, decimals, and percents.

Item responses were coded as correct, incorrect, or missing for the 259 students. However, only 136 of the students answered every item on the APT on both occasions, i.e., 123 students omitted or did not reach at least one item on either occasion. In many cases, clusters of items were omitted in the middle of the tests, which implied that students were omitting the groups of items for which they did not know the answers, rather than reaching a time limit for the test. To deal with this problem of missing data, a 15%-missing-data criterion was employed. A student's response protocol was deleted from the data set if the student omitted more than five items (i.e., 15% of 35 items) on either the pretest or the posttest. This resulted in a group of 220 students on which all further analyses were based. For these 220 students, missing data were coded as incorrect on the assumption that the student did not answer the item because he/she did not know the answer and was unwilling to guess.

Analyses

Differences in achievement level estimates. The question of interest with respect to achievement level estimates was whether there were differences in achievement level estimates due to instruction, i.e., were students growing or gaining in achievement levels throughout the course of instruction? Analyses pertinent to this question included comparisons of the frequency distributions of number-correct scores both before and after instruction and a *t* test for the difference between the means of scores on the pretest and the posttest. Comparisons were also made of the distributions of item difficulties for each administration of the APT. The correlation between scores on the pretest and posttest was computed as an indication of the degree to which the scores were linearly related.

Differences in the structure of achievement. A related but less often investigated issue is whether there are differences in the structure of item responses due to instruction. Investigation of this issue involved computing and comparing the values of coefficient alpha as an index of internal consistency, which is related to the average level of intercorrelation of the items. More germane to this issue, however, was whether the factor structure underlying the test changed with instruction or whether it remained constant. Consequently,

principal axes factor analyses were performed separately on the pretest and posttest item responses. Pearson product-moment correlations were computed between pairs of item responses, and the diagonal elements of the interitem correlation matrices were replaced with initial estimates of the communalities of each item, as given by the squared multiple correlation between that item and the other items in the matrix. An iterative procedure for improving these communality estimates was used, successively extracting factors and re-estimating the communalities. This process continued until the difference between two successive communality estimates was negligible (see Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975).

Random sets of item responses were generated by simulating the responses of 220 students to 35 items such that the probability of a correct answer by any simulee to an item was equal to the difficulty (proportion correct) of that item. This was done separately for the pretest and the posttest. Identical procedures as performed for the real data were carried out for intercorrelating the item responses and factoring the resulting matrix. The results of the factor analyses of real and random data were compared to determine the number of "nonrandom" factors existing in the real data.

The final factor solutions for the pretest and the posttest were then compared in terms of numbers of factors extracted and the similarities between them. Factor similarity was evaluated by computing the root-mean-square deviation, the product-moment correlation coefficient, and the coefficient of congruence between the factor loadings of the factors extracted at each test administration (see Harman, 1976, pp. 343-344). These similarity measures were compared with values obtained from the two sets of random data, as recommended by Nesselroade and Baltes (1970).

Results

Differences in Achievement Level Estimates

Total score differences. Frequency distributions of number-correct scores for both administrations of the APT are presented in Appendix Table A; the frequency polygons are displayed in Figure 1. This figure shows that although the distribution of pretest scores was approximately symmetric, the distribution of posttest scores was negatively skewed, indicating the presence of a ceiling effect. Only four students answered all 35 items correctly on the posttest; an additional 77 students (or 35%) incorrectly answered less than four items. The mean score on the pretest was 22.26, the median was 22.74, and the standard deviation was 5.97. For the posttest these statistics were 28.91, 30.10, and 4.88, respectively. A one-tailed t test for the difference between means of dependent groups was calculated to be 18.67, with probability $p < .0001$.

Item difficulties. The differences in raw score distributions observed between pretest and posttest were mirrored in the distributions of item difficulties for the two administrations of the APT, as shown in Table 1. Although the pretest items were, on the average, answered correctly more often than not, nearly a third of them (i.e., 10 of 35) were answered incorrectly by at least half of the students. For the posttest, however, only two of the items were as difficult. In fact, one third of the items (12 of 35) were answered correctly by more than 90% of the students.

Figure 1
Grouped Frequency Distribution of Number-Correct Scores
for APT Pretest and Posttest

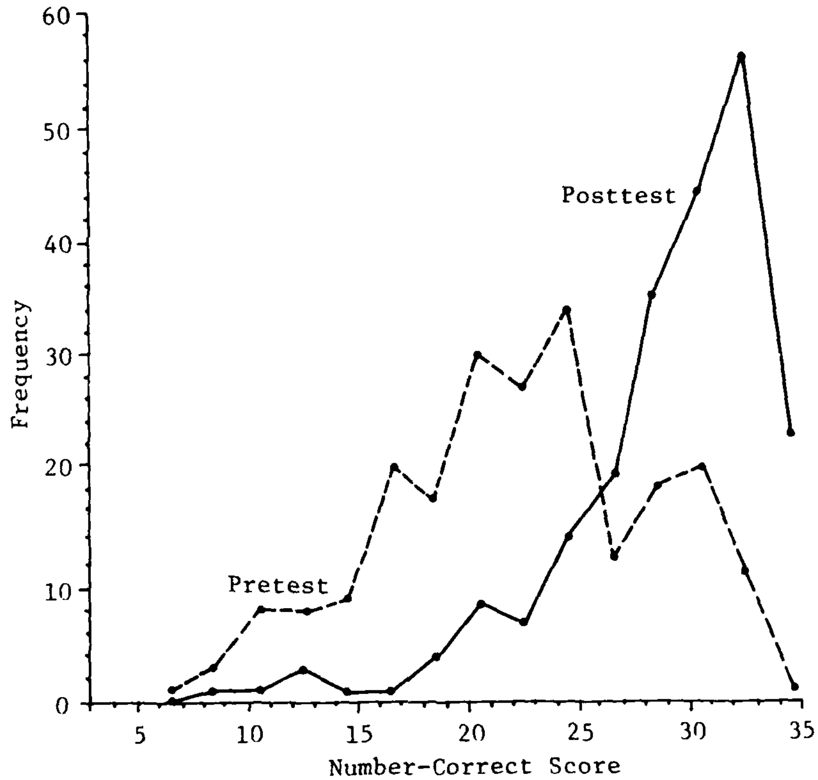


Table 1
Frequency Distributions of
Item Difficulties for APT
Administered as Pretest and as Posttest

Range of Item Difficulty	Number of Items Pretest	Number of Items Posttest
.00 - .10	0	0
.11 - .20	1	0
.21 - .30	1	0
.31 - .40	4	0
.41 - .50	4	2
.51 - .60	5	0
.61 - .70	5	3
.71 - .80	6	9
.81 - .90	5	9
.91 - 1.00	4	12
Mean Difficulty	.64	.83

Correlation between scores. The Pearson product-moment correlation coefficient between number-correct scores at the two administrations of the APT was .542. This relatively low value, coupled with the evidence of mean score increases, reveals that students did not, to a great extent, maintain their relative standings in the course after instruction.

Differences in the Structure of Achievement

Internal consistency reliability. The internal consistency reliability of the APT, as indexed by coefficient alpha, was .836 for the pretest and .835 for the posttest. That the reliability coefficient remained essentially constant provides some evidence for concluding that the items were functioning together in the same manner before and after instruction. However, since the variance of the scores decreased somewhat from pretest to posttest (see Appendix Table A), the stability of coefficient alpha may actually reflect a slight increase in the average interitem correlation.

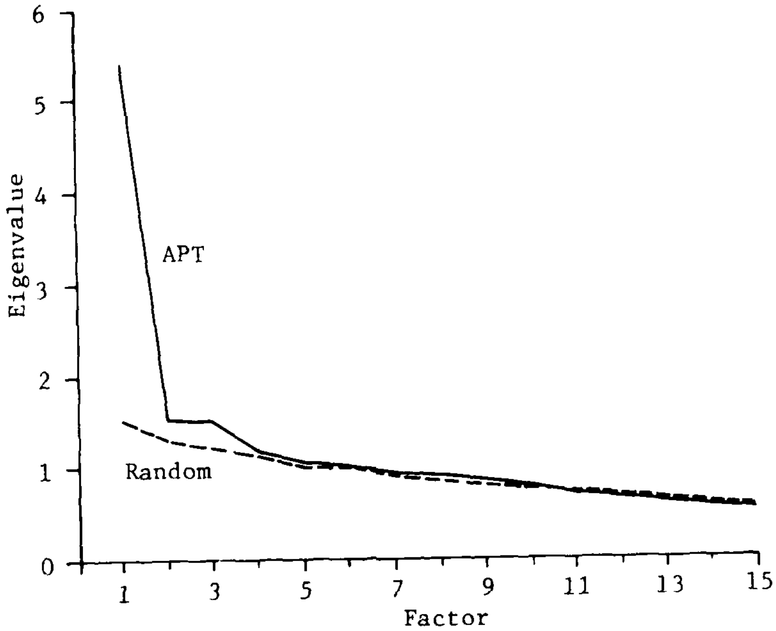
Number of factors extracted. The eigenvalues and percent of total variance accounted for by the first 15 factors from the APT and random data are given in Appendix Table B. The plots of eigenvalues versus factors extracted for both the APT and the random data are given in Figure 2a for the pretest and in Figure 2b for the posttest. In both cases, there was one relatively strong factor in the data; the eigenvalue for the first factor extracted from the APT was much larger than the eigenvalues for the remaining factors in the APT and for all the factors in the random data. The same cannot be said for any of the remaining factors. It was concluded that a one-factor solution adequately described the item response data from both the pretest and the posttest. The FACTOR subroutine in SPSS (Nie et al., 1975) was then run again on the data from each administration, specifying a single-factor solution each time.

Factor similarity. The factor loadings on the single factor extracted from each administration of the APT and from corresponding random data are given in Table 2. The loadings presented in Table 2 were of moderate magnitude; the majority of the loadings were greater than .300, but all were less than .700. The patterns and the magnitudes of the loadings were essentially the same across test administrations. For example, Items 2 through 5 and Item 28 were among the items with the lowest loadings at the pretest; the same was true for these items at the posttest. The items with the highest loadings at the pretest were also among the items with the highest loadings at the posttest. That the magnitude of the loadings was similar for the two administrations can also be seen by comparing the percentage of total variance accounted for by each factor. The single factor extracted from the APT pretest data accounted for 13.92% of the total variance compared to 3.05% for the random data. The factor extracted from the APT posttest data was only slightly stronger, accounting for 14.59% of the total variance as compared to 2.40% in the random data.

Table 3 presents the measures of factor similarity between the APT factor loadings at pretest and at posttest. The root-mean-square deviation between the loadings extracted at each administration is sensitive to differences in the absolute levels of the loadings; low values indicate only minor differences between the values of the two sets of loadings. The root-mean-square deviation was a low .089 for these data. The product-moment correlation coefficient is

Figure 2
Eigenvalues for the First 15 Factors Extracted from the APT and from Corresponding Random Data

(a) Pretest



(b) Posttest

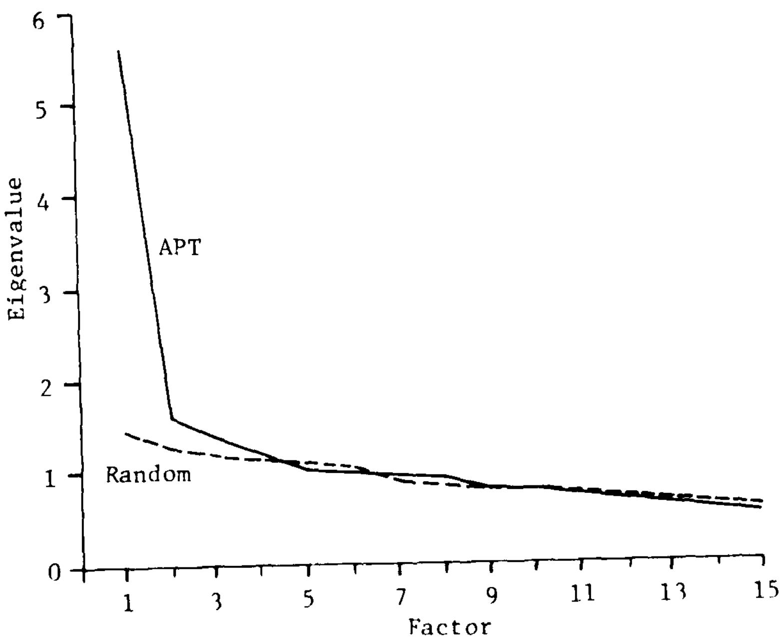


Table 2
Factor Loadings on the Single Factor
Extracted from APT at Pretest and at Posttest,
and from Corresponding Random Data

Item	Pretest APT	Pretest Random Data	Posttest APT	Posttest Random Data
1	.289	.124	.303	-.042
2	.088	.027	-.004	.130
3	.058	.315	.152	-.049
4	.160	.010	.219	-.051
5	.191	.230	.226	.140
6	.263	-.187	.255	.172
7	.332	-.188	.118	.032
8	.315	.147	.383	.036
9	.156	.099	.341	.051
10	.384	.150	.495	-.017
11	.453	-.229	.253	-.277
12	.372	-.178	.244	-.170
13	.255	.007	.259	-.066
14	.394	.345	.338	.136
15	.376	.215	.440	.222
16	.575	-.089	.545	.023
17	.426	.075	.436	-.046
18	.562	-.285	.484	.071
19	.491	-.136	.440	.330
20	.588	.109	.506	.135
21	.580	.029	.676	.025
22	.460	.185	.418	.212
23	.344	-.200	.378	.319
24	.370	.402	.433	.084
25	.338	-.028	.500	.051
26	.460	.108	.560	.005
27	.357	-.074	.467	-.015
28	.117	.044	.141	.054
29	.495	.042	.481	.044
30	.291	.162	.294	.196
31	.292	-.276	.352	.006
32	.378	.018	.386	.017
33	.318	.084	.281	.195
34	.313	.090	.359	.128
35	.339	.153	.267	-.442
Percent of Total Variance	13.92	3.05	14.59	2.40

sensitive only to differences in the patterns of the loadings and was equal to .793. The coefficient of congruence is sensitive to differences in both the level and the pattern of loadings and was a high .972. High values for these latter two indices indicate a high degree of similarity between the two sets of factor loadings. The three figures computed from the parallel random data were

.219, .067, and .118, respectively. It was concluded that the factors extracted from each administration of the APT were nearly identical, both in nature and in strength.

Table 3
Measures of Factor Similarity Between Factor Loadings of APT at Pretest and at Posttest and Between Factor Loadings for Corresponding Random Data

Similarity Index	APT	Random Data
Root-Mean-Square-Deviation	.089	.219
Pearson Product-Moment Correlation	.793	.067
Coefficient of Congruence	.972	.118

Conclusions

Differences in Achievement Level Estimates

There was evidence in these data to conclude that there were gains in mean achievement levels observed after a course of instruction. The difference between the means of scores on the 35-item pretest and posttest was nearly 7 items; the frequency distribution of number-correct scores changed from a symmetric distribution to one that was negatively skewed and displaced to the right. This same effect was mirrored in the distributions of item difficulties. The correlation between the two sets of number-correct scores was .542, indicating that students did not generally maintain their relative standings in the course after instruction. It is not known to what extent this correlation was attenuated due to the ceiling effect observed for the posttest scores.

Differences in the Structure of Achievement

Although there was definitive evidence of mean quantitative change from pretest to posttest, there was no evidence of qualitative differences in the factor structure underlying the item responses. The internal consistency reliability of the test remained constant across administrations. When factor analyses were performed separately on the pretest and posttest interitem correlation matrices, essentially the same factor was extracted each time, as evidenced by the similarity in the levels and pattern of factor loadings.

These data indicate, then, that students in the General College arithmetic classes were indeed leaving the course with increased levels of the same variable measured prior to instruction. The change that occurred within the quarter was quantitative, not qualitative.

STUDY II

Method

Subjects

Data were collected from students enrolled in a general biology class at the University of Minnesota during winter quarter of 1980. A paper-and-pencil pretest was administered to all students present on the first day of class. Computer-administered conventional posttests were given before classroom mid-quarter and final examinations to volunteer students who were awarded extra-credit points for their participation.

Design

Tests. There were two different tests administered at various times throughout the quarter. Test A included 14 items from each of the three content areas covered in class lectures before the midquarter exam (chemistry, the cell, and energy). Test B included 14 items from each of the last three content areas in the course (genetics, reproduction/embryology, and ecology).

Experimental groups. The data collection design for this study is shown in Figure 3. Students were randomly assigned to two experimental groups, Groups 1 and 2, corresponding to the groups of students who were administered one of two pretests--Tests A or B, respectively--on the first day of class. Group 3 included students who were absent for the first class meeting or who did not record on their answer sheet which test they took.

Figure 3
Data Collection Design for Study II

	Group 1	Group 2	Group 3
Pretest	Test A: Content Areas 1-3	Test B: Content Areas 4-6	Other
MQ Posttest	Test A	Test A	Test A
Final Exam Posttest	Test A	Test B	Test B

During the two weeks immediately preceding the classroom midquarter examination, volunteer students were administered conventional tests on the computer (MQ posttest). All these students were administered Test A. During the two weeks immediately preceding the final exam, volunteer students were administered conventional tests on the computer (final exam posttest). Students in Group 1 were readministered Test A; students in Groups 2 and 3 were administered Test B.

All item responses were coded as correct, incorrect, or missing. Missing or omitted items did not present an important problem for this set of data. Nevertheless, the same 15%-missing-data criterion was used here as was used in the previous study: a student's response protocol was deleted from the data set if the student omitted more than 6 (i.e., 15% of 42) items on any one test. For the students included in the analysis, all missing data were coded as incorrect.

Analyses

Differences in achievement level estimates: Test A. The question of whether or not students' achievement level estimates on Test A increased from the pretest to the MQ posttest could be answered by examining the performance of Group 1 students on Test A at both testing occasions. However, the number of students who took Test A both times was small (N = 102) compared to the total number of students who took Test A at the pretest only (N = 276) and the total number of students who took Test A at the MQ posttest only (N = 302). A more powerful test of the difference in mean achievement levels could be performed by combining the data from all students who took Test A at the MQ posttest and by comparing their performance with that of all the students who took Test A as a pretest.

For this comparison, it was necessary to assume that the three groups of students being combined at the MQ posttest were equivalent. Group 1 students were administered Test A both at the pretest and at the MQ posttest. (Although Test A was also administered again at the final exam posttest, the number of Group 1 students who returned to take Test A at the final exam posttest was too small for meaningful comparisons to be made. Hence, Test A analyses were confined to the pretest and MQ posttest administrations.) Performance of Group 1 students on Test A at the MQ posttest can be attributed to the students' underlying ability, to the classroom instruction, and/or to the repetition of items from one occasion to the next. Group 2 students, on the other hand, were administered Test B as the pretest and were administered Test A for the first time at the MQ posttest. Performance of Group 2 students on Test A, then, could be attributed only to the students' underlying ability and/or to the classroom instruction. For some Group 3 students (those who were absent on the first day of class), performance on Test A could also be attributed to their underlying ability and/or to the classroom instruction only. For the other Group 3 students (those who did not record which pretest they took), however, Test A performance could be attributed to their underlying ability, to the classroom instruction, and/or to item repetition. Since these two subgroups of Group 3 students could not be identified and separated for analysis, however, Group 3 was omitted from the following comparison for Test A.

Because students were randomly assigned to Groups 1 and 2 on the first day of class, and because classroom instruction was the same for all students, any differences observed between Groups 1 and 2 on their performance on Test A would

reflect a repetition-of-items effect. If mean test scores of Groups 1 and 2 were not significantly different from each other, then Groups 1 and 2 could be combined at the MQ posttest and compared with all students from Group 1 at the pretest. If a significant repetition-of-items effect were found, then subsequent analyses should be performed only on the data from those students in Group 1. Differences between the scores of Group 1 and Group 2 students were evaluated by the use of a $\underline{t}$ test for the difference between two independent groups and by the Kolmogorov-Smirnov two-sample test for the difference between two frequency distributions.

Analyses relevant to the issue of differences in achievement scores included examination of the frequency distributions and summary statistics of number-correct scores and the distributions of item difficulties from the pretest and the MQ posttest.

Differences in the structure of achievement: Test A. The question of whether or not there were qualitative changes in the nature of achievement test scores due to instruction was again investigated, as in Study I, by analysis of internal consistency reliability coefficients and by separate principal-axes factor analyses. These analyses were performed separately on the pretest and MQ posttest data interitem correlation matrices, with communalities estimated using an iterative procedure, as described in Study I. The number of nonrandom factors was again determined by comparing the results of the factor analyses of Test A data with the results of factor analyses of random data based on items of similar difficulty.

The results of the final solutions from the pretest and the MQ posttest were then compared in terms of the numbers of factors extracted and the similarity of these factors. As in Study I, factor similarity was indexed by the root-mean-square deviation, the product-moment correlation coefficient, and the coefficient of congruence between the factor loadings obtained at each occasion in comparison with values obtained from two sets of random data.

Differences in achievement level estimates: Test B. The question of whether or not students' achievement level estimates on Test B increased from the pretest to the final exam posttest could be answered by examining the performance of Group 2 students on Test B at both testing occasions. However, if no significant repetition-of-items effect was found for Test A (as discussed above), the assumption could be made that there would be no repetition-of-items effect for Test B; then there would be justification for combining the data on Test B from Groups 2 and 3 at the final exam in order to conduct a more powerful test of the difference between mean achievement level estimates. Analyses relevant to this question included examination of the frequency distributions and summary statistics of number-correct scores, and the distributions of item difficulties from the pretest and the final exam posttest.

Differences in the structure of achievement: Test B. As described above, the internal consistency reliability coefficient (coefficient alpha) was computed for Test B at the pretest and at the final exam posttest. Separate principal axes factor analyses were also performed on the Test B data and on parallel random data. The final factor solutions of Test B from the pretest and the final exam posttest were also compared in terms of the number of factors extracted and the similarity of these factors, as was done in Study I and for Test A in this study.

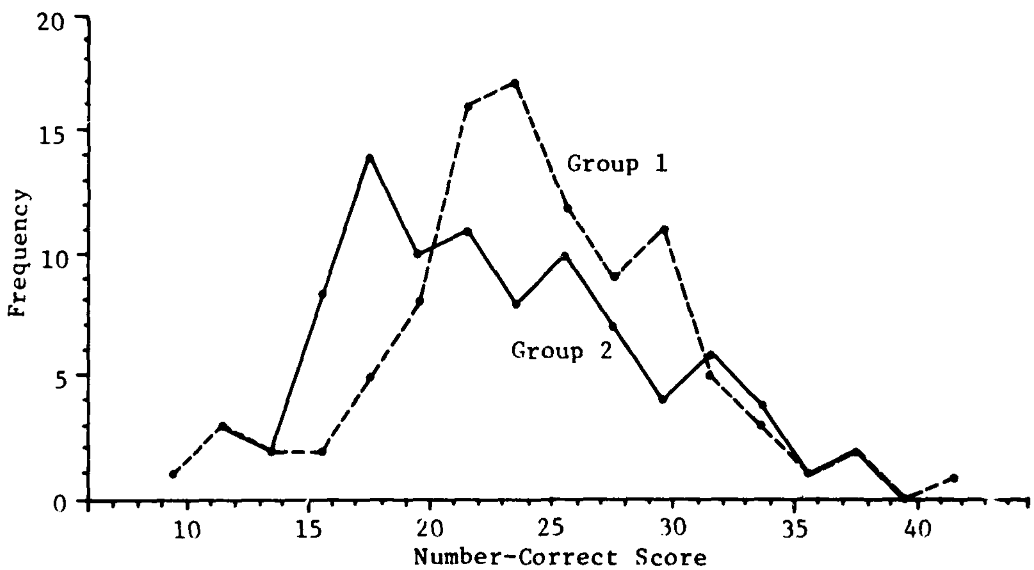
Results

Effect of Item Repetition

The effect on achievement level estimates of repeating items from the pretest to a posttest was evaluated by comparing the performance of students in Groups 1 and 2 on Test A administered before the midquarter exam (MQ posttest). There were 102 students from Group 1 who volunteered to take the MQ posttest, of which 98 met the 15%-missing-data criterion and were retained for analyses. For Group 2 these figures were 101 and 91, respectively.

Appendix Table C presents the frequency distributions of number-correct scores for Test A administered at the MQ posttest to students from Groups 1 and 2; the frequency polygons are displayed in Figure 4. For Group 1 the mean test score was 24.19, the median was 23.79, and the standard deviation was 5.87. For Group 2 these statistics were 22.59, 21.80, and 6.26, respectively. A *t* test of the difference between the means of independent groups was calculated to be 1.98; this was not statistically significant at $p = .01$. The entire frequency distributions of Groups 1 and 2 were compared by using a Kolmogorov-Smirnov two-sample test; the statistic calculated was equal to 7.86, which was not statistically significant at $p = .01$.

Figure 4
Grouped Frequency Distributions of Number-Correct Scores
for Biology Test A Administered at MQ Posttest
for Groups 1 and 2



Although the observed differences were in the predicted direction, the effect of item repetition was not statistically significant. Hence, the question of identifying and separating the two subgroups of Group 3 was no longer relevant, and the Test A MQ posttest scores of students in Groups 1, 2, and 3 were combined for comparison with the scores of all students who took Test A on the first day of class. Since some of the students who took the test at the pretest

did not take it at the posttest, the correlation between scores at pretest and posttest was not computed.

Missing Data

There were 276 students who were administered Test A at the pretest; of these 272 met the 15%-missing-data criterion and were retained for further analyses. The combined total of students who took Test A at the MQ posttest was 302, and 283 of these were retained for further analyses.

Because there was no effect of item repetition observed for Test A, the performance of Group 2 students who were administered Test B at the pretest was compared with the performance of students from both Groups 2 and 3 who were administered Test B at the final exam posttest. There were 283 students who were administered Test B at the pretest, of which 277 met the 15%-missing-data criterion and were retained for further analyses. A total of 169 students took Test B at the final exam posttest, and 163 of them were retained for further analyses.

Differences in Achievement Level Estimates: Test A

Total score differences. Frequency distributions of number-correct scores on Test A at both testing occasions are presented in Appendix Table D; the frequency polygons appear in Figure 5. Both distributions are approximately symmetric, with the distribution of MQ posttest scores displaced to the right. The mean of the pretest scores was 15.97, with a standard deviation of 3.97. For the MQ posttest scores, these figures were 23.46 and 5.99, respectively. The mean score difference between the two occasions was 7.49. Because there was some overlap between the students in the two groups, the groups were not strictly independent, nor were they strictly dependent. A t test for the difference between two independent means, although technically inappropriate, would yield a conservative test of the significance of this difference. This test resulted in t (df = 553) = 17.34, $p < .001$.

Item difficulties. The frequency distributions of item difficulties for Test A at both testing occasions are given in Table 4. As indicated earlier, the pretest was somewhat difficult: 74% of the items were answered correctly by less than half the students, and no item was answered correctly more than 80% of the time. After instruction, more than half the items (23 of 42) were answered correctly by 51% to 90% of the students, although five items were answered correctly less than 30% of the time.

Differences in the Structure of Achievement: Test A

Internal consistency reliability. Coefficient alpha for Test A when administered on the first day of class was .490. This low value indicates that the average interitem correlation was correspondingly small. After instruction, coefficient alpha increased to .787 for the same set of items. Although this value is not high for a 42-item test, it represents a substantial increase over the value obtained at the pretest. The difference between these two figures may indicate that the items were functioning as a set differently after instruction than they were before instruction and/or it may reflect the increase in the variance of the number-correct scores.

Figure 5
Grouped Frequency Distributions of Number-Correct Scores
for Biology Test A Administered at Pretest and at MQ Posttest

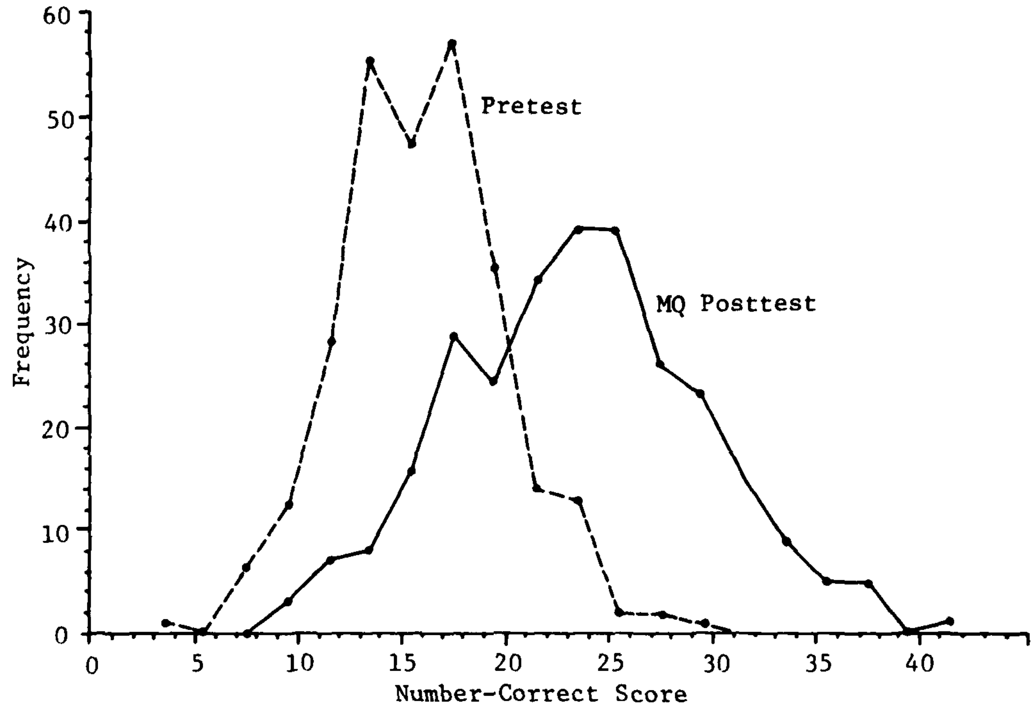
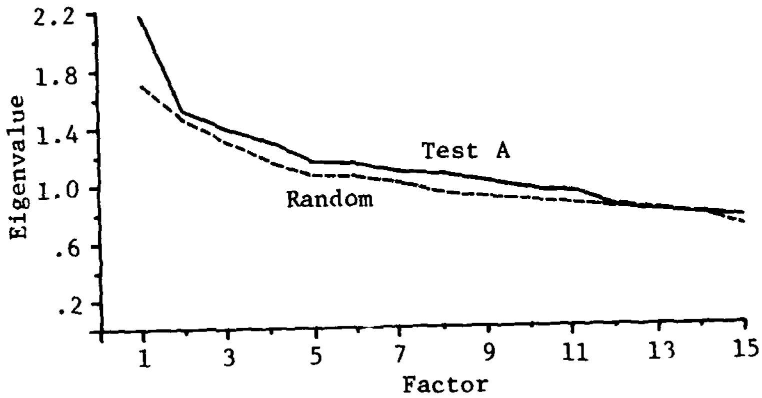


Table 4
Frequency Distributions of Item
Difficulties for Biology Test A
Administered at Pretest
and at MQ Posttest

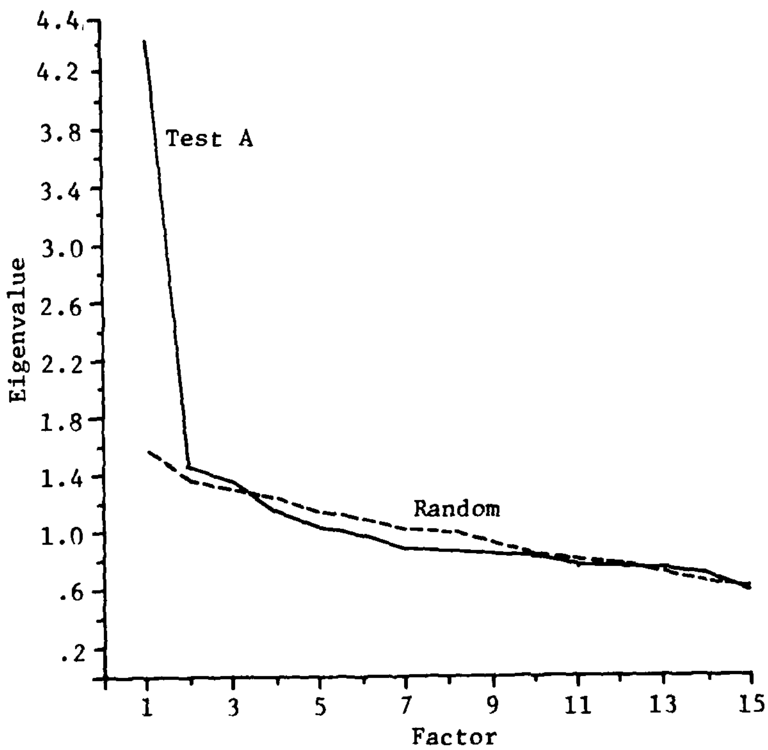
Range of Item Difficulty	Number of Items	
	Pretest	Posttest
.00 - .10	1	1
.11 - .20	8	1
.21 - .30	8	3
.31 - .40	9	7
.41 - .50	5	7
.51 - .60	4	5
.61 - .70	2	5
.71 - .80	5	5
.81 - .90	0	8
.91 - 1.00	0	0
Mean Difficulty	.38	.56

Figure 6
Eigenvalues for the First 15 Factors Extracted from Test A
Administered at Pretest and at MQ Posttest,
and from Corresponding Random Data

(a) Pretest



(b) MQ Posttest



Number of factors extracted. Appendix Table E presents the eigenvalues and percent of total variance accounted for by the first 15 factors from Test A and from corresponding random data. Figure 6a presents the plots of eigenvalues versus factors extracted from Test A and from random data at the pretest, and Figure 6b presents results for the MQ posttest. Comparison of the results from Test A with the results from the corresponding random data revealed that there was one weak factor present in the pretest and one stronger factor present in the posttest.

Factor similarity. Table 5 presents the factor loadings on the single factor extracted at each testing occasion from Test A and from corresponding random data. Comparison of these factor loadings reveals that the loadings from the MQ posttest were, in general, higher than those from the pretest. No loading from the pretest was greater than .391, and nearly two-thirds of the factor loadings (26 of 42) were less than .200. For the MQ posttest, the highest loading was .502, but 81% of the factor loadings (34 of 42) were greater than .200.

This result can also be seen by comparing the percentages of total variance accounted for by the single factor at each administration. For the pretest that figure was 3.96% (as compared to 2.88% for the random data); for the MQ posttest the factor accounted for 9.36% of the total variance (as compared to 2.79% for the random data). Both of these percentages are small for a 42-item test, indicating that the factor was relatively weak, even at the MQ posttest.

The pattern of factor loadings did not appear to be consistent across test administrations. The items with the lowest loadings at the pretest did not emerge as the items with the lowest loadings at the MQ posttest, and the same was true for the items with the highest loadings.

Table 6 presents the measures of factor similarity between the two sets of loadings for Test A and the corresponding random data. The root-mean-square deviation between the two sets of loadings for Test A, sensitive to differences in levels of the loadings, was .195, a high value when considered in conjunction with the relatively narrow range of loadings observed in these data. The product-moment correlation coefficient between the loadings, sensitive to pattern differences, was a low .373. The coefficient of congruence was .780. The similarity measures obtained from the random data were .160, .549, and .548, respectively. All these figures reveal that the factors extracted from Test A on the two occasions were not substantially more similar than were factors extracted from randomly generated data.

These data reveal, then, that the factor extracted from Test A at the pretest differed substantially from that extracted at the MQ posttest. Although there was a sizeable increase in the number-correct scores after instruction, there was a corresponding change in the first factor underlying the item responses. This indicates that the pretest and the MQ posttest measured quite different variables, even though they were composed of exactly the same items.

Differences in Achievement Level Estimates: Test B

Total score differences. Frequency distributions of number-correct scores on Test B at both testing occasions are given in Appendix Table F; their frequency polygons are presented in Figure 7. The distribution of final exam post-

Table 5
Factor Loadings on the Single Factor
Extracted from Biology Test A at Pretest and at MQ Posttest,
and from Corresponding Random Data

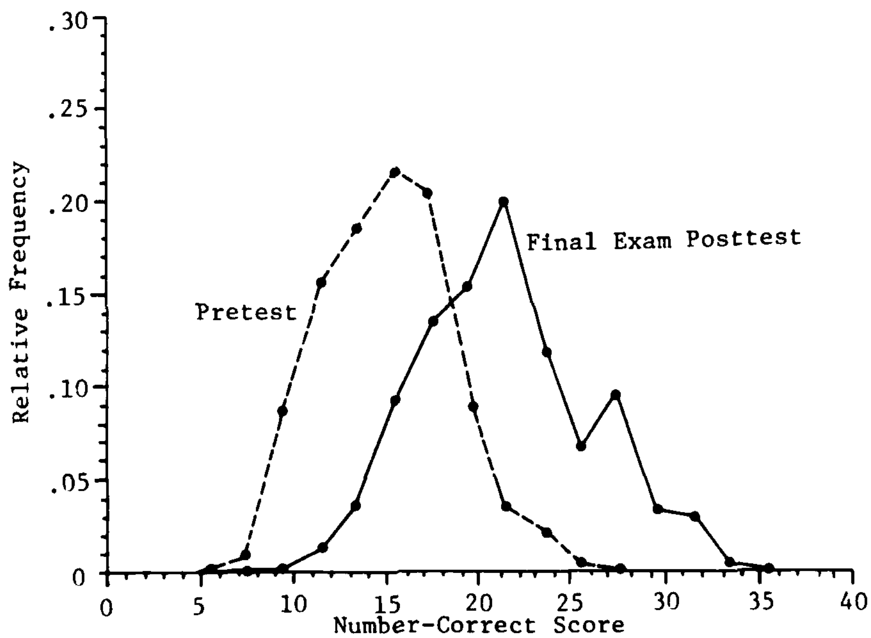
Item	Pretest		Posttest	
	Test A	Random Data	Test B	Random Data
1	.068	-.032	.186	.158
2	.024	-.026	.133	-.205
3	.331	-.245	.161	.051
4	.115	.163	.279	.150
5	-.002	-.238	.276	-.099
6	.206	-.054	.008	.029
7	.280	.191	.372	.121
8	.191	-.246	.333	-.153
9	.272	.096	.408	.120
10	.027	-.005	.367	-.002
11	.291	-.163	.154	-.154
12	.103	-.035	.207	.011
13	.370	.327	.502	.208
14	.391	-.197	.344	-.223
15	.042	.440	.388	.418
16	.273	-.010	.341	.296
17	.133	-.042	.335	.079
18	.239	-.105	.310	-.162
19	.388	.021	.276	.162
20	.205	.362	.410	.222
21	.115	-.059	.316	-.098
22	.223	-.040	.479	-.161
23	.383	.060	.298	.024
24	.245	.067	.373	-.114
25	.052	-.053	.228	.187
26	-.024	-.116	.246	-.105
27	.039	.091	.478	.083
28	.015	-.094	.143	.060
29	.117	.061	.315	.244
30	.343	-.139	.372	-.224
31	.095	.070	.200	.057
32	.194	-.027	.284	-.154
33	.043	.179	.272	.255
34	.059	-.050	.249	.337
35	.096	-.150	.301	.190
36	-.026	.148	.245	.206
37	.221	-.139	.340	-.021
38	.107	-.185	.227	-.095
39	.106	.282	.241	-.016
40	-.111	-.344	-.030	.077
41	-.124	.162	.164	-.041
42	.063	.113	.422	.117
Percent of Total Variance	3.96	2.88	9.36	2.79

Table 6
Measures of Factor Similarity Between Factor Loadings for Test A at Pretest and at MQ Posttest, and Between Factor Loadings from Corresponding Random Data

Similarity Index	Test A	Random Data
Root-Mean-Square-Deviation	.195	.160
Pearson Product-Moment Correlation	.373	.549
Coefficient of Congruence	.780	.548

test scores is approximately symmetric, while that of the pretest scores is slightly positively skewed. The mean of the pretest scores was 15.18, with standard deviation 3.54. For the final exam posttest scores, these figures were 21.47 and 4.58, respectively. The score difference between the mean scores on the two occasions was 6.29. As before, a t test for the difference between two independent means, though technically inappropriate, was conducted as a conservative test of this difference; here, t (df = 438) = 16.15, p < .001.

Figure 7
Grouped Relative Frequency Distributions of Number-Correct Scores for Biology Test B Administered at Pretest and at Final Exam Posttest



Item difficulties. The frequency distributions of item difficulties for Test B at both testing occasions are given in Table 7. As was observed for the

number-correct scores, the pattern of item difficulties reveals that the pretest was somewhat difficult: 74% of the items were answered correctly by less than half the students, and only two items were answered correctly more than 80% of the time. At the end of the course, more than half the items (22 of 42) were answered correctly by the majority of students, although 12 items were answered correctly less than 30% of the time.

Table 7
Frequency Distributions of Item Difficulties for Biology Test B Administered at Pretest and at Final Exam Posttest

Range of Item Difficulty	Number of Items	
	Pretest	Posttest
.00 - .10	4	2
.11 - .20	9	3
.21 - .30	8	7
.31 - .40	3	4
.41 - .50	7	4
.51 - .60	5	2
.61 - .70	2	10
.71 - .80	2	5
.81 - .90	2	4
.91 -1.00	0	1
Mean Difficulty	.36	.51

Differences in the Structure of Achievement: Test B

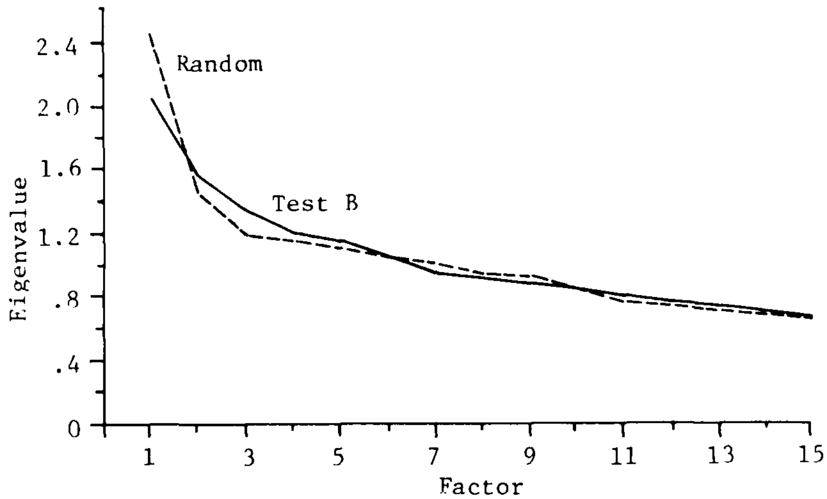
Internal consistency reliability. When administered at the pretest on the first day of class, coefficient alpha for Test B was .398, increasing to .630 when administered at the final exam posttest. These low values indicate that the average interitem correlation coefficient was correspondingly small. Even though both reliability coefficients were relatively low, the fact that the reliability coefficient increased from .40 to .63 may be an indication that the items were functioning as a set differently after instruction than they were before instruction. As before, however, this increase may simply be reflecting the increase in the variance of the test scores.

Number of factors extracted. Appendix Table G presents the eigenvalues and percentages of total variance accounted for by the first 15 factors extracted from Test B and from corresponding random data. Figure 8a presents the plots of these eigenvalues versus factors extracted at the pretest, and Figure 8b presents similar data from the final exam posttest. Comparison of the results from the real data with the results from the random data reveals that there was no factor stronger than one extracted from the random data in the pretest, but one stronger factor was extracted from Test B at the final exam posttest.

Factor similarity. Table 8 presents the factor loadings on the single factor extracted at each testing occasion from Test B and from corresponding random data. Comparison of these factor loadings reveals that the loadings from the

Figure 8
Eigenvalues for the First 15 Factors Extracted from Biology Test B
Administered at Pretest and at Final Exam Posttest,
and from Corresponding Random Data

(a) Pretest



(b) Final Exam Posttest

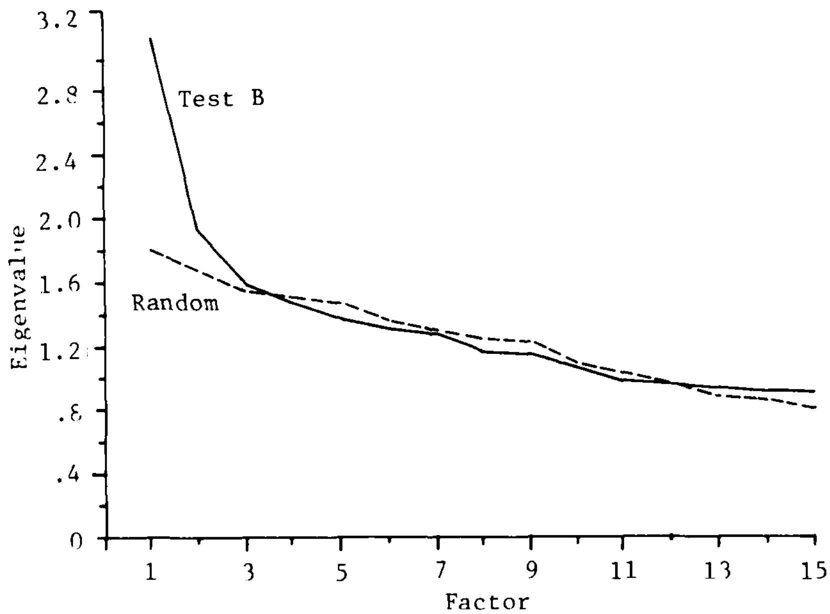


Table 8
Factor Loadings on the Single Factor Extracted from Biology Test B at Pretest and at Final Exam Posttest and from Corresponding Random Data

Item	Pretest		Posttest	
	Test B	Random Data	Test B	Random Data
1	.131	.088	.295	-.044
2	.073	.087	.310	.377
3	-.023	-.168	.193	.258
4	.218	.122	.416	.098
5	.252	-.286	.137	.113
6	.268	.145	.240	.179
7	.191	.145	.256	-.236
8	.127	-.113	.296	.246
9	-.044	.293	.273	-.066
10	.323	-.320	.255	.296
11	.193	.471	.202	.060
12	.164	.117	.311	-.239
13	.393	-.111	.371	.161
14	-.007	-.136	.438	.030
15	.228	-.085	.261	.045
16	.329	-.099	.301	.284
17	.246	-.252	.310	.193
18	.154	.381	.372	-.073
19	.192	-.098	.241	.006
20	-.027	.341	.193	-.013
21	.231	-.151	.307	.092
22	-.239	-.156	.268	.411
23	.459	.213	.299	.162
24	.062	.067	.079	.140
25	.009	.182	.330	-.037
26	.045	-.101	.174	-.044
27	-.101	.034	-.112	-.057
28	.130	-.080	.043	.112
29	.296	-.245	.084	.088
30	.215	.077	.155	.328
31	.252	.179	.397	.003
32	.278	.020	.177	-.123
33	-.045	.045	-.112	-.082
34	.028	-.277	.137	.003
35	.012	.384	.165	.093
36	.166	-.012	-.071	.047
37	-.115	-.034	-.023	-.026
38	.018	.060	-.002	.009
39	.082	.120	.011	.053
40	.040	.109	.178	-.088
41	.013	-.457	.205	-.015
42	-.058	.510	-.111	-.071
Percent of Total Variance	3.69	4.70	5.96	2.54

final exam posttest were, in general, slightly higher than those from the pretest. The highest pretest loading was .459, and nearly two-thirds of the factor loadings (27 of 42) were less than .200. For the final exam posttest, the highest loading was .438, but more than half of the factor loadings (23 of 42) were greater than .200.

This result can also be seen by comparing the percentage of total variance accounted for by the single factor extracted at each administration. For the pretest, that figure was 3.69% (as compared to 4.70% accounted for by the random factor); for the final exam posttest, the factor accounted for 5.96% of the total variance (as compared to 2.54% for the random data). Both of these percentages are very small, indicating that the factor was relatively weak.

The pattern of factor loadings did not appear consistent across test administrations. The items with the lowest loadings at the pretest did not necessarily emerge as the items with the lowest loadings at the final exam posttest, and the same was true for the items with the highest loadings.

Table 9 presents the measures of factor similarity for Test B. The root-mean-square deviation between the two sets of loadings for Test B, sensitive to differences in levels of the loadings, was .177, a high value when considered in conjunction with the relatively narrow range of loadings observed in this data but lower than the .300 observed for the two sets of random data. The product-moment correlation coefficient between the loadings, sensitive to pattern differences, was a low .399 as contrasted with $\underline{r} = -.327$ for the random data. The coefficient of congruence was .697 for Test $\overline{B}$ and -.255 for the random data. Although the comparison of the similarity measures reveals that the factor loadings for Test B were more congruent than the corresponding sets of random data, the degree of similarity was so low that these factors could not justifiably be considered congruent.

Table 9
Measures of Factor Similarity Between Factor Loadings from Test B at Pretest and at Final Exam Posttest, and Between Factor Loadings from Corresponding Random Data

Similarity Index	Test B	Random Data
Root-Mean-Square Deviation	.177	.300
Pearson Product-Moment Correlation	.399	-.327
Coefficient of Congruence	.696	-.255

These data reveal, then, that the factor extracted from Test B at the pretest differed from the factor extracted at posttest. As was observed for Test A, there was a sizeable increase in the number-correct scores, accompanied by a change in the factor underlying the item responses. This indicates that the pretest and the final exam posttest were measuring quite different variables, even though they were composed of exactly the same items.

Conclusions

Differences in Achievement Level Estimates

The results from both Test A and Test B indicate that there were mean differences in achievement level estimates (number-correct scores) that accompanied classroom instruction. On the average, test scores increased after relevant course instruction; for these data, scores increased between 6 and 7.5 points on a 42-item test. The increases in these test scores were not attributable to the effect of item repetition. Although the differences were in the predicted direction, neither a t test nor the Kolmogorov-Smirnov two-sample test were significant at $p = .01$.

Differences in the Structure of Achievement

There were substantial differences in the structure of item responses to the items on both biology tests--Test A and Test B--from the pretest to the posttest. Large increases in the internal consistency reliability coefficient may reflect corresponding changes in the average interitem correlation coefficients. That is, changes in the way the items functioned together as a set were evident after instruction took place. This same effect was observed when the factor structures of the tests at both administrations were compared. Although only one factor was extracted at each administration of each test, the factor at each pretest was very weak and bore little relationship to the factor extracted later in the course, as reflected in the patterns and levels of the factor loadings.

DISCUSSION AND CONCLUSIONS

The results of these studies show that the use of simple difference scores to measure change in classroom achievement may not be appropriate for all subject matter areas. The use of simple difference scores, or some derivative thereof, assumes that there is only a quantitative difference between pretest and posttest achievement levels due to a course of instruction. That is, the assumption is made that a pretest measures a baseline amount of some knowledge or trait and that classroom instruction results in increased levels of the same trait, as indicated by higher scores on the same, or a similar, test.

This assumption was supported by the results of the mathematics data. There was a large and statistically significant difference observed in achievement test scores obtained before and after instruction. That the same trait was being measured both times was indicated by the high degree of similarity of the underlying factor structure of the test when examined at both points in time. The only change observed in the mathematics test scores was, then, a quantitative one, reflected in increases in mean number-correct score after classroom instruction in mathematics.

The results were quite different for the two biology tests examined. Factor analyses of the pretests revealed the presence of one very weak factor for each pretest. One slightly stronger factor also emerged at each of the posttests, but there was very little correspondence between the pretest and posttest

factors. Even though mean test scores increased after instruction, there was a corresponding difference in the factors underlying test performance. The change that occurred in the biology test scores, then, was a qualitative one, where the tests were measuring different variables before and after instruction. Evaluating gains in achievement by computing pretest-posttest difference scores cannot be justified under these circumstances.

That the results from these two studies are different has important bearing on the issue of program evaluation and the measurement of change. The question of whether the difference in test scores that follows classroom instruction or program participation is quantitative or qualitative must be answered before any attempt at quantifying change can legitimately be made. For some courses of instruction, the application of classical change-score methodology may be defended on the grounds that the only change observed was quantitative; for others, the use of such methodology may not be justified. Clearly, further research is needed to define those areas where the use of change scores or their derivatives may be warranted.

REFERENCES

Anastasi, A. The influence of specific experience upon mental organization. Genetic Psychology Monographs, 1936, 18, 245-355.

Cronbach, L. J., & Furby, L. How should we measure "change"--or should we? Psychological Bulletin, 1970, 74, 68-80.

Ferguson, G. A. On learning and human ability. Canadian Journal of Psychology, 1954, 8, 95-112.

Fleishman, E. A. A factor analysis of intra-task performance on two psychomotor tasks. Psychometrika, 1953, 18, 45-55.

Fleishman, E. A. A comparative study of aptitude patterns in unskilled and skilled psychomotor performances. Journal of Applied Psychology, 1957, 41, 263-272.

Fleishman, E. A. Abilities at different stages of practice in rotary pursuit performance. Journal of Experimental Psychology, 1960, 60, 162-171.

Fleishman, E. A., & Fruchter, B. Factor structure and predictability of successive stages of learning Morse code. Journal of Applied Psychology, 1960, 44, 97-101.

Fleishman, E. A., & Hempel, W. E., Jr. Changes in factor structure of a complex psychomotor task as a function of practice. Psychometrika, 1954, 19, 239-252.

Fleishman, E. A., & Hempel, W. E., Jr. The relationship between abilities and improvement with practice in a visual discrimination reaction task. Journal of Experimental Psychology, 1955, 49, 301-312.

Games, P. A. A factorial analysis of verbal learning tasks. Journal of Experimental Psychology, 1962, 63, 1-11.

Garrett, H. E. A developmental theory of intelligence. American Psychologist, 1946, 1, 372-378.

Greene, E. B. An analysis of random and systematic changes with practice. Psychometrika, 1943, 8, 37-52.

Harman, H. H. Modern factor analysis (3rd ed.). Chicago: University of Chicago Press, 1976.

Kingsbury, G. G., & Weiss, D. J. Effect of point-in-time in instruction on the measurement of achievement (Research Report 79-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, August 1979.

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin, 1963.

Nesselroade, J. R., & Baltes, P. B. On a dilemma of comparative factor analysis: A study of factor matching based on random data. Educational and Psychological Measurement, 1970, 30, 935-948.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill, 1975.

Querishi, M. Y. Patterns of psycholinguistic development during early and middle childhood. Educational and Psychological Measurement, 1967, 27, 353-365.

Reinert, G. Comparative factor analytic studies of intelligence throughout the human life span. In L. R. Goulet & P. B. Baltes (Eds.), Life-span developmental psychology: Research and theory. New York: Academic Press, 1970.

Sullivan, J. P., & Moran, L. J. Association structures of bright children at age six. Child Development, 1967, 38, 793-800.

Swartz, J. D., & Moran, L. J. Association structures of bright children at ages nine and twelve. Multivariate Behavioral Research, 1968, 3, 189-198.

Wohlwill, J. F. Methodology and research strategy in the study of developmental change. In L. R. Goulet & P. B. Baltes (Eds.), Life-span developmental psychology: Research and theory. New York: Academic Press, 1970.

Woodrow, H. The relation between abilities and improvement with practice. Journal of Educational Psychology, 1938, 29, 215-230.

Woodrow, H. The application of factor-analysis to problems of practice. Journal of General Psychology, 1939, 21, 457-460. (a)

Woodrow, H. Factors in improvement with practice. Journal of Psychology, 1939, 7, 55-70. (b)

Woodrow, H. The relation of verbal ability to improvement with practice in verbal tests. Journal of Educational Psychology, 1939, 30, 179-186. (c)

Appendix: Supplementary Tables

Table A
Frequency Distributions of Number-Correct Scores
for APT Pretest and Posttest (N=220)

	Pretest			Posttest		
Score	Frequency	Percent	Cumulative Percent	Frequency	Percent	Cumulative Percent
35	0	0	100.0	4	1.8	100.0
34	1	.5	100.0	20	9.1	98.2
33	4	1.8	99.5	28	12.7	89.1
32	7	3.2	97.7	29	13.2	76.4
31	7	3.2	94.5	19	8.6	63.2
30	13	5.9	91.4	25	11.4	54.5
29	5	2.3	85.5	16	7.3	43.2
28	13	5.9	83.2	19	8.6	35.9
27	5	2.3	77.3	11	5.0	27.3
26	8	3.6	75.0	8	3.6	22.3
25	14	6.4	71.4	7	3.2	18.6
24	20	9.1	65.0	7	3.2	15.5
23	17	7.7	55.9	6	2.7	12.3
22	10	4.5	48.2	1	0.5	9.5
21	14	6.4	43.6	5	2.3	9.1
20	16	7.3	37.3	4	1.8	6.8
19	6	2.7	30.0	1	0.5	5.0
18	11	5.0	27.3	3	1.4	4.5
17	11	5.0	22.3	1	0.5	3.2
16	9	4.1	17.3	0	0.0	2.7
15	7	3.2	13.2	0	0.0	2.7
14	2	0.9	10.0	1	0.5	2.7
13	4	1.8	9.1	2	0.9	2.3
12	4	1.8	7.3	1	0.5	1.4
11	7	3.2	5.5	0	0.0	0.9
10	1	0.5	2.3	1	0.5	0.9
9	0	0.0	1.8	0	0.0	0.5
8	3	1.4	1.8	1	0.5	0.5
7	1	0.5	0.5	0	0.0	0.0
Mean	22.26			28.91		
SD	5.97			4.88		
Median	22.74			30.10		
Mode	24			32		

Table B
Eigenvalues and Percent of Total Variance
Accounted for by First 15 Factors Extracted from the APT
at Pretest and at Posttest, and from Corresponding Random Data

	Pretest				Posttest			
	APT		Random Data		APT		Random Data	
Factor	Eigen-Value	% Total Variance	Eigen-Value	% Total Variance	Eigen-Value	% Total Variance	Eigen-Value	% Total Variance
1	5.350	15.3	1.545	4.4	5.590	16.0	1.419	4.1
2	1.555	4.4	1.308	3.7	1.605	4.6	1.253	3.6
3	1.539	4.4	1.229	3.5	1.337	3.8	1.161	3.3
4	1.209	3.5	1.139	3.3	1.171	3.3	1.134	3.2
5	1.086	3.1	1.029	2.9	1.034	3.0	1.052	3.0
6	1.016	2.9	.993	2.8	1.006	2.9	1.023	2.9
7	.942	2.7	.890	2.5	.986	2.8	.896	2.6
8	.892	2.5	.865	2.5	.939	2.7	.828	2.4
9	.876	2.5	.822	2.3	.839	2.4	.814	2.3
10	.794	2.3	.767	2.2	.797	2.3	.790	2.3
11	.739	2.1	.745	2.1	.756	2.2	.770	2.2
12	.666	1.9	.692	2.0	.675	1.9	.732	2.1
13	.607	1.7	.634	1.8	.660	1.9	.702	2.0
14	.597	1.7	.600	1.7	.604	1.7	.666	1.9
15	.553	1.6	.566	1.6	.533	1.5	.600	1.7

Table C
Frequency Distribution of Number-Correct Scores for
Biology Test A at MQ Posttest for Students in Groups 1 and 2

	Group 1 (N=98)			Group 2 (N=91)		
Score	Frequency	Percent	Cumulative Percent	Frequency	Percent	Cumulative Percent
41	1	1.0	100.0	0	0.0	100.0
40	0	0.0	99.0	0	0.0	100.0
39	0	0.0	99.0	0	0.0	100.0
38	0	0.0	99.0	1	1.1	100.0
37	2	2.0	99.0	1	1.1	98.9
36	1	1.0	96.9	0	0.0	97.8
35	0	0.0	95.9	1	1.1	97.8
34	1	1.0	95.9	3	3.3	96.7
33	2	2.0	94.9	1	1.1	93.4
32	3	3.1	92.9	2	2.2	92.3
31	2	2.0	89.8	4	4.4	90.1
30	5	5.1	87.8	1	1.1	85.7
29	6	6.1	82.7	3	3.3	84.6
28	4	4.1	76.5	1	1.1	81.3
27	5	5.1	72.4	6	6.6	80.2
26	6	6.1	67.3	5	5.5	73.6
25	6	6.1	61.2	5	5.5	68.1
24	7	7.1	55.1	2	2.2	62.6
23	10	10.2	48.0	6	6.6	60.4
22	7	7.1	37.8	5	5.5	53.8
21	9	9.2	30.6	6	6.6	48.4
20	3	3.1	21.4	6	6.6	41.8
19	5	5.1	18.4	4	4.4	35.2
18	2	2.0	13.3	9	9.9	30.8
17	3	3.1	11.2	5	5.5	20.9
16	1	1.0	8.2	5	5.5	15.4
15	1	1.0	7.1	3	3.3	9.9
14	1	1.0	6.1	1	1.1	6.6
13	1	1.0	5.1	1	1.1	5.5
12	2	2.0	4.1	1	1.1	4.4
11	1	1.0	2.0	2	2.2	3.3
10	0	0.0	1.0	1	1.1	1.1
9	1	1.0	1.0	0	0	0.0
Mean	24.19			22.59		
SD	5.87			6.26		
Median	23.79			21.80		
Mode	23			18		

Table D
Frequency Distribution of Number-Correct Scores
for Biology Test A at Pretest and at MQ Posttest

	Pretest (N=272)			Posttest (N=283)		
Score	Frequency	Percent	Cumulative Percent	Frequency	Percent	Cumulative Percent
41	0	0.0	100.0	1	0.4	100.0
40	0	0.0	100.0	0	0.0	99.6
39	0	0.0	100.0	0	0.0	99.6
38	0	0.0	100.0	1	0.4	99.6
37	0	0.0	100.0	4	1.4	99.3
36	0	0.0	100.0	2	0.7	97.9
35	0	0.0	100.0	3	1.1	97.2
34	0	0.0	100.0	4	1.4	96.1
33	0	0.0	100.0	5	1.8	94.7
32	0	0.0	100.0	6	2.1	92.9
31	0	0.0	100.0	9	3.2	90.8
30	1	0.0	100.0	8	2.8	87.6
29	0	0.0	99.6	15	5.3	84.8
28	1	0.4	99.6	9	3.2	79.5
27	1	0.4	99.3	17	6.0	76.3
26	0	0.0	98.9	16	5.7	70.3
25	2	0.7	98.9	23	8.1	64.7
24	5	1.8	98.2	15	5.3	56.5
23	8	2.9	96.3	24	8.5	51.2
22	6	2.2	93.4	15	5.3	42.8
21	8	2.9	91.2	19	6.7	37.5
20	9	3.3	88.2	14	4.9	30.7
19	25	9.2	84.9	10	3.5	25.8
18	23	8.5	75.7	16	5.7	22.3
17	34	12.5	67.3	13	4.6	16.6
16	23	8.5	54.8	9	3.2	12.0
15	24	8.8	46.3	7	2.5	8.8
14	30	11.0	37.5	5	1.8	6.4
13	25	9.2	26.5	3	1.1	4.6
12	13	4.8	17.3	3	1.1	3.5
11	15	5.5	12.5	4	1.4	2.5
10	7	2.6	7.0	1	0.4	1.1
9	5	1.8	4.4	2	0.7	0.7
8	3	1.1	2.6	0	0.0	0.0
7	3	1.1	1.5	0	0.0	0.0
6	0	0.0	0.4	0	0.0	0.0
5	0	0.0	0.4	0	0.0	0.0
4	1	0.4	0.4	0	0.0	0.0
Mean	15.97			23.46		
SD	3.97			5.99		
Median	15.94			23.35		
Mode	17			23		

Table E
Eigenvalues and Percent of Total Variance Accounted for by First 15 Factors Extracted from Biology Test A at Pretest and at MQ Posttest and Corresponding Random Data

	Pretest				MQ Posttest			
	Test A		Random Data		Test A		Random Data	
Factor	Eigen-Value	% Total Variance	Eigen-Value	% Total Variance	Eigen-Value	% Total Variance	Eigen-Value	% Total Variance
1	2.200	5.2	1.706	4.1	4.411	10.5	1.572	3.7
2	1.512	3.6	1.456	3.5	1.440	3.4	1.358	3.2
3	1.395	3.3	1.299	3.1	1.349	3.2	1.302	3.1
4	1.298	3.1	1.172	2.8	1.167	2.8	1.238	2.9
5	1.167	2.8	1.053	2.5	1.026	2.4	1.134	2.7
6	1.136	2.7	1.044	2.5	.980	2.3	1.103	2.6
7	1.075	2.6	1.001	2.4	.895	2.1	1.017	2.4
8	1.064	2.5	.913	2.2	.885	2.1	.999	2.4
9	1.004	2.4	.901	2.1	.844	2.0	.915	2.2
10	.951	2.3	.876	2.1	.825	2.0	.839	2.0
11	.923	2.2	.845	2.0	.784	1.9	.810	1.9
12	.820	2.0	.813	1.9	.771	1.8	.783	1.9
13	.805	1.9	.793	1.9	.748	1.8	.726	1.7
14	.757	1.8	.751	1.8	.696	1.7	.663	1.6
15	.726	1.7	.677	1.6	.598	1.4	.611	1.5

Table F
Frequency Distribution of Number-Correct Scores
for Biology Test B at Pretest and at Final Exam Posttest

	Pretest (N=277)			Posttest (N=163)		
Score	Frequency	Percent	Cumulative Percent	Frequency	Percent	Cumulative Percent
33	0	0.0	100.0	1	0.6	100.0
32	1	0.4	100.0	2	1.2	99.4
31	0	0.4	100.0	3	1.8	98.2
30	0	0.4	100.0	1	0.6	96.3
29	0	0.4	100.0	5	3.1	95.7
28	0	0.4	100.0	8	4.9	92.6
27	0	0.4	100.0	8	4.9	87.7
26	0	0.0	100.0	6	3.7	82.8
25	1	0.4	99.6	5	3.1	79.1
24	2	0.7	99.3	8	4.9	76.1
23	4	1.4	98.6	13	8.0	71.2
22	4	1.4	97.1	16	9.8	63.2
21	6	2.2	95.7	17	10.4	53.4
20	10	3.6	93.5	15	9.2	42.9
19	12	4.3	89.9	10	6.1	33.7
18	27	9.7	85.6	12	7.4	27.6
17	31	11.2	75.8	10	6.1	20.2
16	29	10.5	64.6	10	6.1	14.1
15	30	10.8	54.2	5	3.1	8.0
14	29	10.5	43.3	3	1.8	4.9
13	23	8.3	32.9	3	1.8	3.1
12	22	7.9	24.5	0	0.0	1.2
11	21	7.6	16.6	2	1.2	1.2
10	16	5.8	9.0	0	0.0	0.0
9	7	2.5	3.2	0	0.0	0.0
8	2	0.7	0.7	0	0.0	0.0
Mean	15.18			21.47		
SD	3.54			4.58		
Median	15.12			21.18		
Mode	17			21		

Table G
Eigenvalues and Percent of Total Variance Accounted for by First 15 Factors Extracted from Biology Test B at Pretest and at Final Exam Posttest and from Corresponding Random Data

	Pretest				Final Exam Posttest			
	Test B		Random Data		Test B		Random Data	
Factor	Eigen-Value	% Total Variance	Eigen-Value	% Total Variance	Eigen-Value	% Total Variance	Eigen-Value	% Total Variance
1	2.043	4.9	2.440	5.8	3.124	7.4	1.810	4.3
2	1.551	3.7	1.448	3.4	1.920	4.6	1.678	4.0
3	1.345	3.2	1.190	2.8	1.590	3.8	1.550	3.7
4	1.204	2.9	1.146	2.7	1.480	3.5	1.513	3.6
5	1.152	2.7	1.098	2.7	1.383	3.3	1.466	3.5
6	1.065	2.5	1.053	2.5	1.309	3.1	1.370	3.3
7	.932	2.2	.999	2.4	1.284	3.1	1.305	3.1
8	.911	2.2	.929	2.2	1.167	2.8	1.234	2.9
9	.887	2.1	.920	2.2	1.151	2.7	1.215	2.9
10	.835	2.0	.852	2.0	1.059	2.5	1.105	2.6
11	.796	1.9	.770	1.8	.978	2.3	1.030	2.5
12	.781	1.9	.739	1.8	.964	2.3	.966	2.3
13	.747	1.8	.702	1.7	.927	2.2	.895	2.1
14	.709	1.7	.684	1.6	.911	2.2	.857	2.0
15	.685	1.6	.668	1.6	.819	2.0	.803	1.9

Distribution List

Navy

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Meryl S. Baker
NPRDC
Code P[illegible]
San Diego, CA 92152

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Chief of Naval Education and Training
Liason Office
Air Force Human Resource Laboratory
Flying Training Division
WILLIAMS AFB, AZ 85224

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Patrick R. Harrison
Psychology Course Director
LEADERSHIP & LAW DEPT. (7b)
DIV. OF PROFESSIONAL DEVELOPMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr William Montague
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. William Nordbrock
Instructional Program Development
Bldg. 90
NET-PDCD
Great Lakes Naval Training Cent
IL 60088

1 Ted M. I. Yellen
Technical Information Office
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

1 Office of Naval Research
Code 441
800 N. Quincy Street
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Office of the Chief of Naval Operations
Research Development & Studies Branch
(OP-115)
Washington, DC 20350

1 LT Frank C. Petho, MSC, USN (Ph.D)
Selection and Training Research Div
Human Performance Sciences Dept.
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 Dr. Bernard Rimland [illegible]
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Worth Scanland, Director
Research, Development, [illegible]
N-5
Naval Education and Training Command
NAS, Pensacola, FL [illegible]

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group
(TAEG)
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

1 Roger Weissinger-Baylon
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr John H. Wolfe
Code P310
U. S. Navy Personnel Research and
Development Center
San Diego, CA 92152

Army

1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 DR. JAMES L. RANEY
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Commandant
US Army Institute of Administration
Attn: Dr. Sherrill
FT Benjamin Harrison, IN 46256

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

1 U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Dr. Alfred R. Fregly
AFOSR/NL, Bldg. 410
Bolling AFB
Washington, DC 20332

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 David R. Hunter
AFHRL/MOAM
Brooks AFB, TX 78235

1 Research and Measurment Division
Research Branch, AFMPC/MPCYPR
Randolph AFB, TX 78148

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

2 3700 TCHTW/TTGH Stop 32
Sheppard AFB, TX 76311

1 Dr. Joe Ward, Jr.
AFHRL/MPMD
Brooks AFB, TX 78235

Marines

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower Utilization
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 Headquarters, U. S. Marine Corps
Code MPI-20
Washington, DC 20380

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 Major Michael L. Patrow, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Chief, Psychological Reserch Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Other DoD

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

1 Director, Research and Data
OASD(MRA&L)
3B919, The Pentagon
Washington, DC 20301

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Dr. Wayne Sellman
Office of the Assistant Secretary
of Defense (MRA & L)
2B269 The Pentagon
Washington, DC 20301

1 DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civil Govt

1 Mr. Richard McKillip
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Service
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. James Algina
University of Florida
Gainesville, FL 32611

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 Dr. John Annett
Department of Psychology
University of Warwick
Coventry CV4 7AL
ENGLAND

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Capt. J. Jean Belanger
Training Development Division
Canadian Forces Training System
CFTSHQ, CFB Trenton
Astra, Ontario KOK 1B0

1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
Israel

1 Dr. Werner Birke
DezWPs im Streitkraefteamt
Postfach 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. R. Darrel Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Liaison Scientists
Office of Naval Research,
Branch Office , London
Box 39 FPO New York 09510

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

1 DR. JOHN F. BROCK
Honeywell Systems & Research Center
(MN 17-2318)
2600 Ridgeway Parkway
Minneapolis, MN 55413

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. William E. Coffman
Director, Iowa Testing Programs
334 Lindquist Center
University of Iowa
Iowa City, IA 52242

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036

1 Dr. Hans Crombag
Education Research Center
University of Leyden
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

1 Dr. Fritz Drasgow
Yale School of Organization and Management
Yale University
Box 1A
New Haven, CT 06520

1 Mike Durmeyer
Instructional Program Development
Building 90
NET-PDCD
Great Lakes NTC, IL 60088

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Benjamin A. Fairbank, Jr.
McFann-Gray & Associates, Inc.
5825 Callaghan
Suite 225
San Antonio, Texas 78228

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Daniel Gopher
Industrial & Management Engineering
Technion-Israel Institute of Technology
Haifa
ISRAEL

1 Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

1 Dr. Ron Hambleton
School of Education
University of Massechusetts
Amherst, MA 01002

1 Dr. Delwyn Harnisch
University of Illinois
242b Education
Urbana, IL 61801

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Dustin H. Houston
Wicat, Inc.
Box 986
Orem, UT 84057

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Jack Hunter
2122 Coolidge St.
Lansing, MI 48906

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Professor John A. Keats
University of Newcastle
AUSTRALIA 2308

1 Mr. Jeff Kelety
Department of Instructional Technology
University of Southern California
Los Angeles, CA 92007

1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

1 Dr. Marcy Lansman
Department of Psychology, NI 25
University of Washington
Seattle, WA 98195

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
Netherlands

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Dr. Erik McWilliams
Science Education Div. and Research
National Science Foundation
Washington, DC 20550

1 Professor Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army Navy Drive
Arlington, VA 22202

1 Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Cirle, NW
Washington, DC 20036

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 Dr. Steven E. Poltrock
Department of Psychology
University of Denver
Denver,CO 80208

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MINRAT M. L. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 1328
D-53 BONN 1, GERMANY

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
4 Hill Hall
Columbia, MO 65211

1 [illegible]
[illegible]
[illegible]
[illegible]

1 [illegible]
[illegible]
[illegible]
[illegible]

1 Dr. Lawrence Rudner
[illegible] Avenue
Takoma Park, MD [illegible]

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 Frank L. Schmidt
Department of Psychology
Bldg. GG
George Washington University
Washington, DC 20052

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Thomas G. Sticht
Director, Basic Skills Division
HUMRRO
300 N. Washington Street
Alexandria,VA 22314

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Brad Sympson
Psychometric Research Group
Educational Testing Service
Princeton, NJ 08541

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. David Vale
Assessment Systems Corporation
2395 University Avenue
Suite 306
St. Paul, MN 55114

1 Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08540

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2
WEST GERMANY

PREVIOUS PUBLICATIONS (CONTINUED)

77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977.
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977.
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977.
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977.
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977.
Final Report: Computerized Ability Testing, 1972-1975. April 1976.
76-5. Effects of Item Characteristics on Test Fairness. December 1976.
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976.
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976.
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976.
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975.
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975.
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975.
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975.
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975.
74-5. Strategies of Adaptive Ability Measurement. December 1974.
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974.
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974.
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974.
73-4. An Empirical Study of Computer-Administered Two-Stage Ability Testing. October 1973.
73-3. The Stratified Adaptive Computerized Ability Test. September 1973.
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973.

Copies of these reports are available, while supplies last, from:
Computerized Adaptive Testing Laboratory
N660 Elliott Hall
University of Minnesota
75 East River Road
Minneapolis MN 55455 U.S.A.

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

81-4. Factors Influencing the Psychometric Characteristics of an Adaptive Testing Strategy for Test Batteries. November 1981.
81-3. A Validity Comparison of Adaptive and Conventional Strategies for Mastery Testing. September 1981.
Final Report: Computerized Adaptive Ability Testing. April 1981.
81-2. Effects of Immediate Feedback and Pacing of Item Presentation on Ability Test Performance and Psychological Reactions to Testing. February 1981.
81-1. Review of Test Theory and Methods. January 1981.
80-5. An Alternate-Forms Reliability and Concurrent Validity Comparison of Bayesian Adaptive and Conventional Ability Tests. December 1980.
80-4. A Comparison of Adaptive, Sequential, and Conventional Testing Strategies for Mastery Decisions. November 1980.
80-3. Criterion-Related Validity of Adaptive Testing Strategies. June 1980.
80-2. Interactive Computer Administration of a Spatial Reasoning Test. April 1980.
Final Report: Computerized Adaptive Performance Evaluation. February 1980.
80-1. Effects of Immediate Knowledge of Results on Achievement Test Performance and Test Dimensionality. January 1980.
79-7. The Person Response Curve: Fit of Individuals to Item Characteristic Curve Models. December 1979.
79-6. Efficiency of an Adaptive Inter-Subtest Branching Strategy in the Measurement of Classroom Achievement. November 1979.
79-5. An Adaptive Testing Strategy for Mastery Decisions. September 1979.
79-4. Effect of Point-in-Time in Instruction on the Measurement of Achievement. August 1979.
79-3. Relationships among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods. April 1979.
Final Report: Bias-Free Computerized Testing. March 1979.
79-2. Effects of Computerized Adaptive Testing on Black and White Students. March 1979.
79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979.
78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978.
78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.
78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978.
78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.
78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978.
77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977.

-continued overleaf-

FILMED

3-8